

The American Journal of Human Genetics, Volume 93

Supplemental Data

**RFMix: A Discriminative Modeling Approach
for Rapid and Robust Local-Ancestry Inference**

Brian K. Maples, Simon Gravel, Eimear E. Kenny, and Carlos D. Bustamante

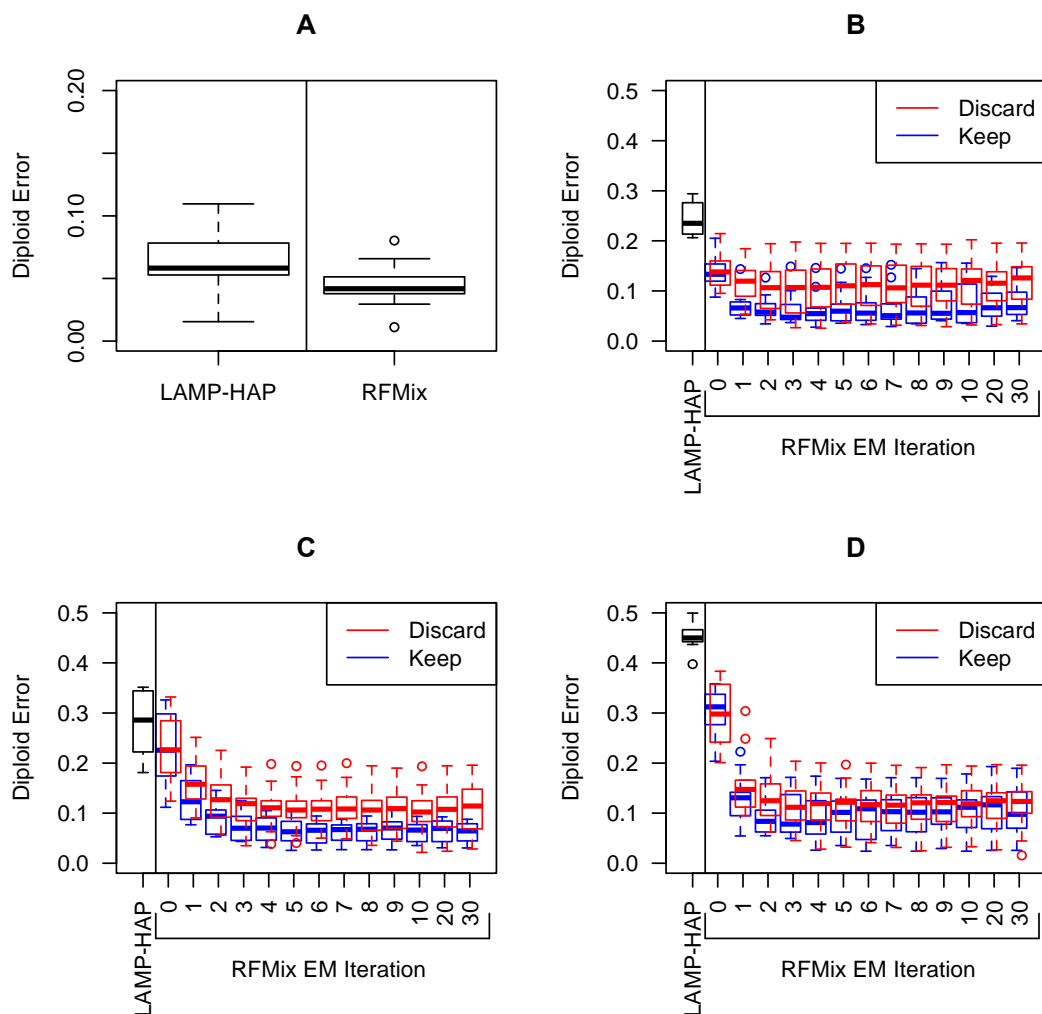


Figure S1. Diploid Ancestry Error when Keeping vs. Discarding the Original Reference Panels after the Initial Inference Step

We simulated 10 Latino individuals as described in the text and used reference panels composed of 30 ideal samples (A). We then simulated an additional 30 Latino individuals and used reference panels composed of (B) 3 ideal samples, (C) 30 proxy samples, or (D) 3 proxy samples. The 0th EM iteration of RFMix refers to the initial round of learning and inference shown in Figure 1. Red boxplots correspond to RFMix runs where the original reference panel was discarded after the 0th EM iteration and blue corresponds to the original configuration.

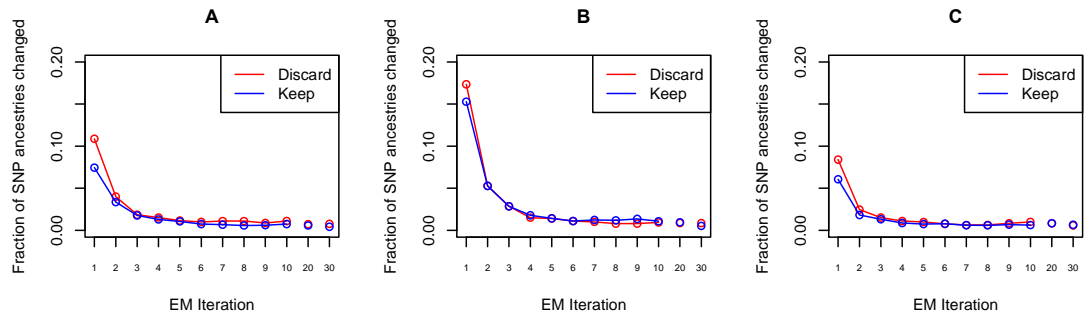


Figure S2. Fraction of Inferred SNP Ancestries Changed at Each EM Iteration

We simulated 10 Latino individuals as described in the text (A) and used reference panels composed of 3 ideal samples, (B) 30 proxy samples, or (C) 3 proxy samples. After the initial inference step, the original reference panels were either kept (blue line) or discarded (red line) and the fraction of SNPs in the admixed samples that changed ancestry assignment were recorded after each EM iteration.

Table S1. Comparison of Accuracies of Nonvoting and Voting Approaches

	<u>Nonvoting</u>	<u>Voting</u>
JPT/CHS	0.88	0.90
GBR/TSI	0.76	0.76
Latinos	0.98	0.98

Haploid accuracy when inferring local ancestry in simulated admixtures using the original approach compared to using a majority vote from overlapping windows at each SNP.